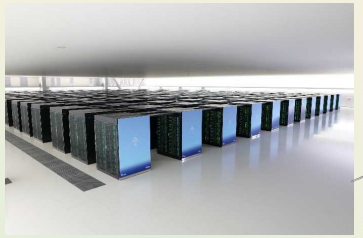
Energy-Efficient Co-Design for Next-Generation supercomputers



富岳
Fugaku

Fumiyoshi Shoji
Operations and Computer Technologies Division
RIKEN Center for Computational Science

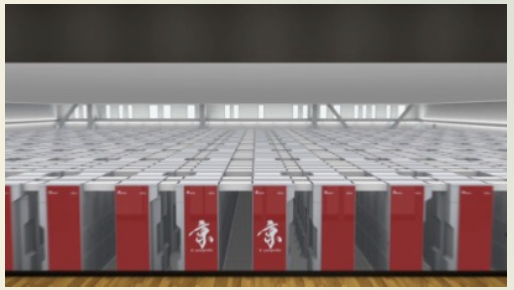
@BoF on HPC Procurement, Energy Efficiency and Sustainability

Evolution of Japanese Flagship Systems

©JAXA

©Tsukuba Univ.

©:JAMSTC



1993: NWT

Numerical Wind Tunnel
Domain: Fluid Dynamics

1996: CP-PACS

Computational Physics
Domain: Computational Physics

2002: Earth Simulator

Global Climate
Domain: Earth Science

2011: K computer

First General-Purpose
Target: All Science Fields

- **Before the K computer, it was a system designed for a specific domain.**
- **The K computer was the first system designed to cover various fields of science and engineering, and collaboration between application scientists and system designers (Co-design) became more important.**
- **Co-design facilitates:**
  - **Sharing Technical and/or Political Constraints**
  - **Finding the Optimal Compromise**
  - **Early Launch of Application Science**

K computer: Dawn of Co-design

Maximizing Scientific Outcomes

In the K computer R&D Project, a co-design process was implemented, in which the development of strategic applications in "**life sciences**" and "**materials science**" proceeded in parallel with system development.

Gordon-Bell prizes as a remarkable outcome:

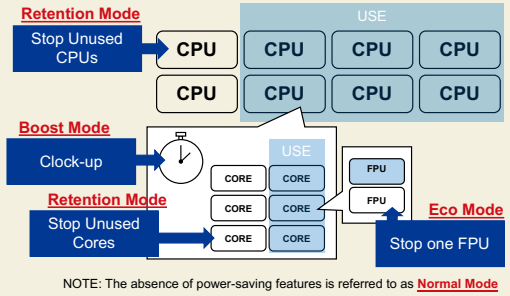
- **2011: "First-principles calculations of electron states of a silicon nanowire with 100,000 atoms on the K computer"**
- **2012: "4.45 Pflops astrophysical N-body simulation on K computer -- The gravitational trillion-body problem"**

Co-design accelerates performance in real-world applications and scientific discovery

Co-design in Fugaku Development: Goals, Implementation, and Achievements

富岳 Fugaku

- **Target-Driven Approach**
  - Focused on solving 9 national priority social and scientific issues.
- **HW/SW Synergy**
  - Co-designed the A64FX processor with ARM SVE and FP16 support for HPC and AI workloads.
- **Power Control Co-design**
  - Developed "Power Knobs" for dynamic hardware and software energy management.



NOTE: The absence of power-saving features is referred to as Normal Mode

Table 6.1: priority issues and their target applications

no.	priority issue	organizations	target application
1	Innovative drug discovery infrastructure through functional control of biomolecular systems	RIKEN Center for Biosystems Dynamics Research and 6 other institutions	GENESIS
2	Integrated computational life science to support personalized and preventive medicine	The Institute of Medical Science, the University of Tokyo and 4 other institutions	Genomon
3	Development of integrated simulation systems for hazards and disasters induced by earthquakes and tsunamis	Institute, the University of Tokyo and 4 other institutions	GAMERA
4	Advancement of meteorological and global environmental predictions utilizing observational " Big Data "	Japan Agency for Marine-Earth Science and Technology (JAMSTEC) and other 5 institutions	NICAM +LETKF
5	Development of new fundamental technologies for high-efficiency energy creation, conversion/storage and use	Institute for Molecular Science, National Institute of Natural Sciences and 8 other institutions	NTChem
6	Accelerated development of innovative clean energy systems	School of Engineering, the University of Tokyo and 11 other institutions	ADVENTURE
7	Creation of new functional devices and high-performance materials to support next-generation industries (CDMSI)	The Institute for Solid State Physics, the University of Tokyo and 9 other institutions	RSDFT
8	Development of innovative design and production processes that lead the way for the manufacturing industry in the near future	Institute of Industrial Science, the University of Tokyo and 7 other institutions	FFB
9	Elucidation of the fundamental laws and evolution of the universe	Center for Computational Sciences, University of Tsukuba and 10 other institutions	LQCD

Figure 6.1: social and scientific priority issues to be tackled by using supercomputer Fugaku

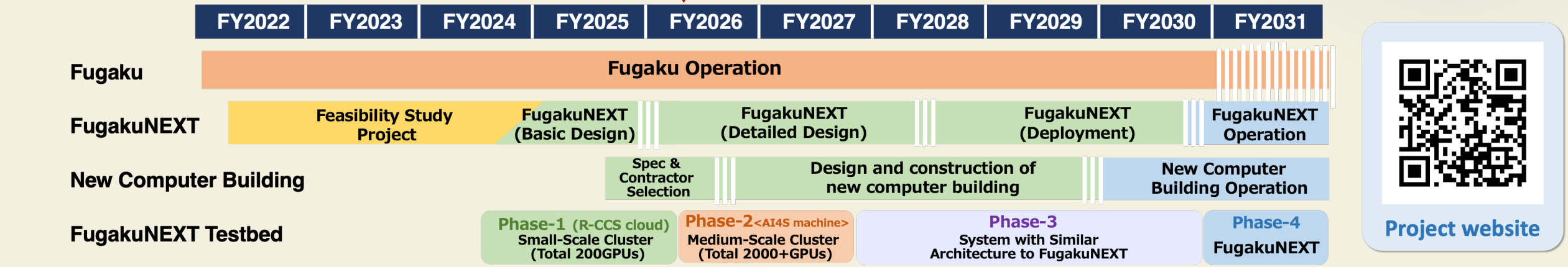
Achievements:

- **Up to 100x improvement in application performance compared to the K computer.**
- **Reduced power consumption(-7%) by incentivizing user cooperation for energy-efficient operation**

https://www.r-ccs.riken.jp/assets/uploads/2022/03/fs2020-report-e.pdf

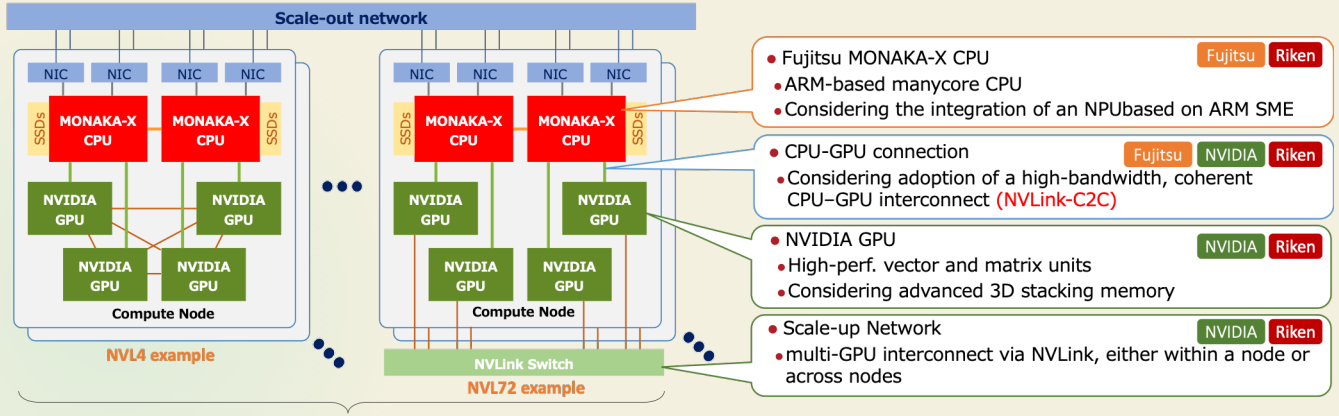
FugakuNEXT project status

	K computer	Fugaku	FugakuNEXT
Launch	Sep. 2012	Mar. 2021	2030
System type	CPU	CPU	CPU+GPU
Peak power consumption(HPL)	12.7MW	29.9MW	<40MW
Average power	10MW	17MW	TBD
Water cooling ratio	70%(CPU/NIC)	90%(CPU/MEM/NIC)	>95%(CPU/GPU/MEM/NIC/PSU)
PUE	1.3-1.4	1.3-1.35	<1.1

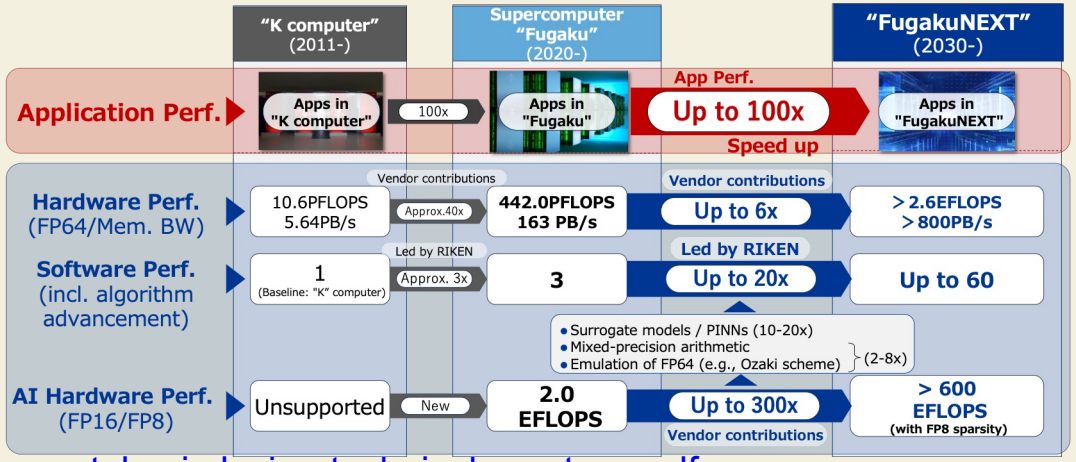


Project website

Overview of FugakuNEXT System Architecture



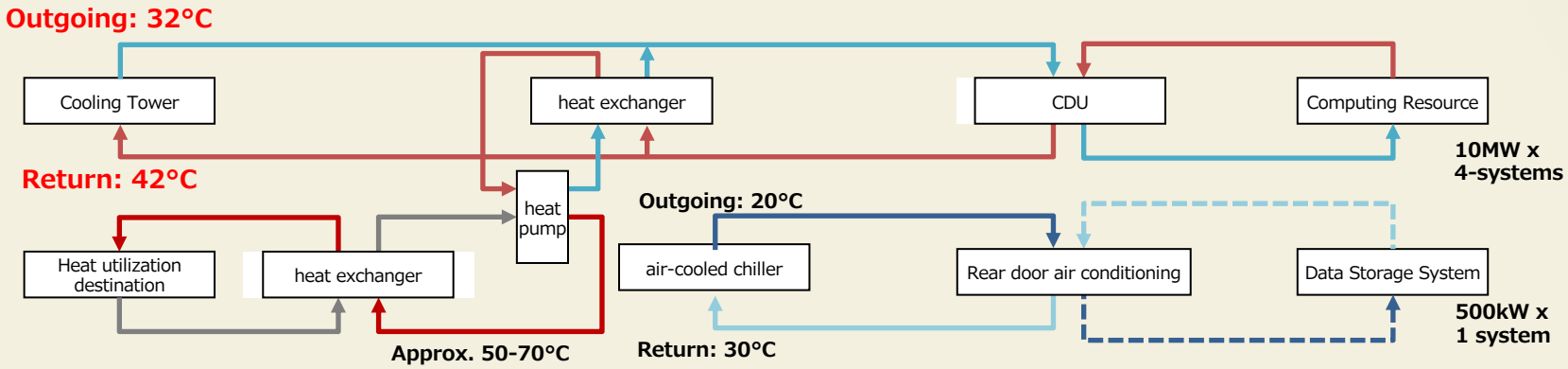
Road to Achieving up to 100x Application Performance



FugakuNEXT basic design technical report: https://www.r-ccs.riken.jp/fugaku-next/fugakunext_basicdesign-technicalreport_en.pdf

Strategy: Energy-Efficiency/Carbon-neutrality

- Warm-water DLC cooling and chiller-less cooling system to reduce energy for cooling
  - There aren't strict restrictions on water consumption.



Peak partial PUE
1.06

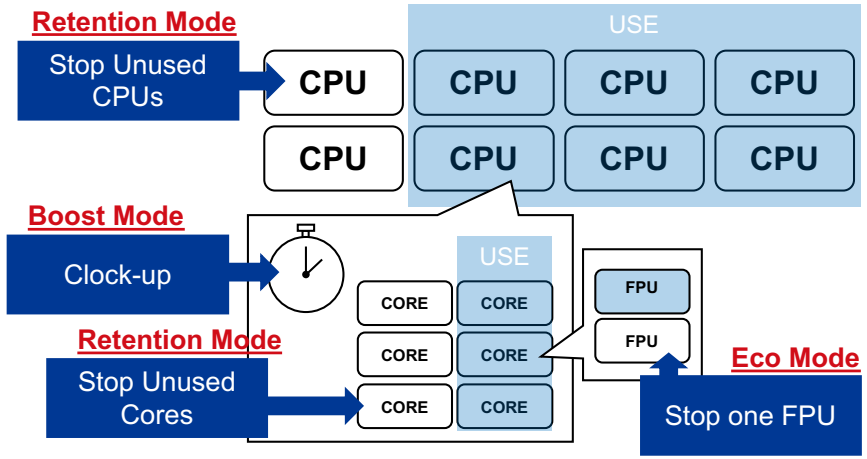
- Automatically sets the optimal power settings for the jobs to reduce energy to solution
  - Analyze the job history data to identify the optimal power settings
- Procure electricity directly from the market and operate in a way that responds to price fluctuations to reduce the cost of electricity (and carbon footprint)
  - The job scheduler controls the mix of jobs (power-consuming jobs and not) in response to the price.
  - An incentive program will serve to strike a balance between user needs and energy-efficient operations.
- Heat-Reuse to reduce carbon footprint
  - It’s not easy for us to find heat consumers around our center.

These will be evaluated in the operation of the FugkauNEXT testbeds

Backup slides

Incentivizing user cooperation for energy-efficient operation

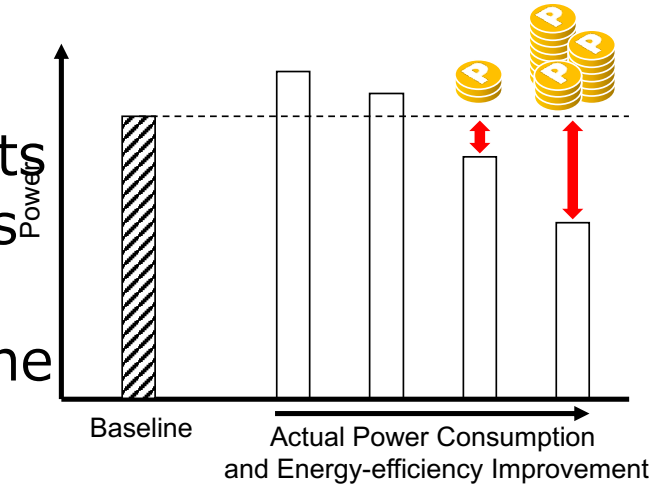
Fugaku equips the power knobs, functions for power-saving



NOTE: The absence of power-saving features is referred to as Normal Mode

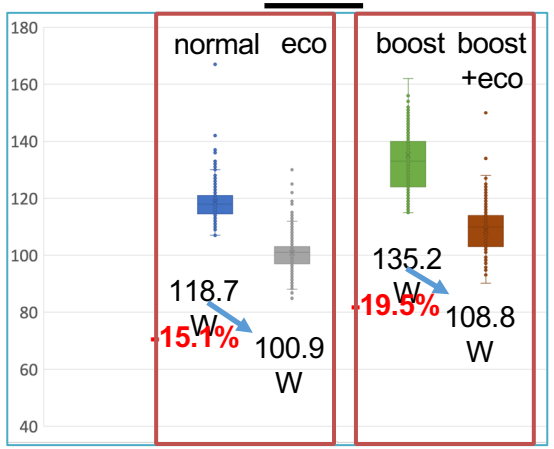
"Fugaku point" to quantify the user cooperation for power saving by user jobs

- Users can earn points when their jobs runs with less power consumption than the standard.

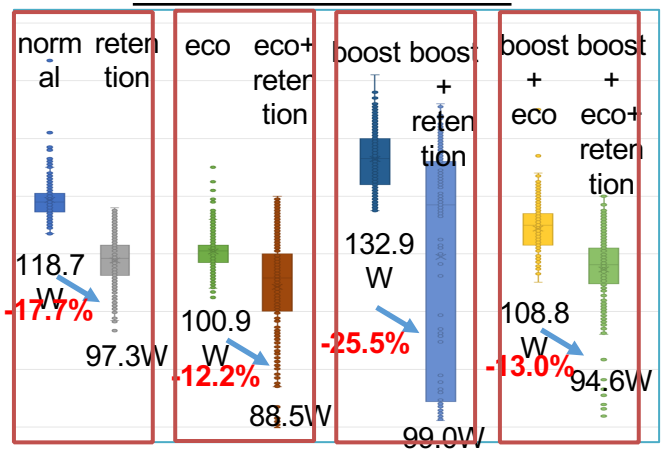


actual power saving

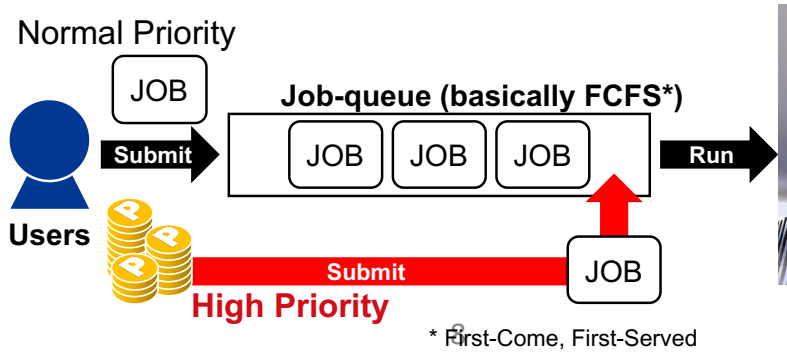
eco



core retention

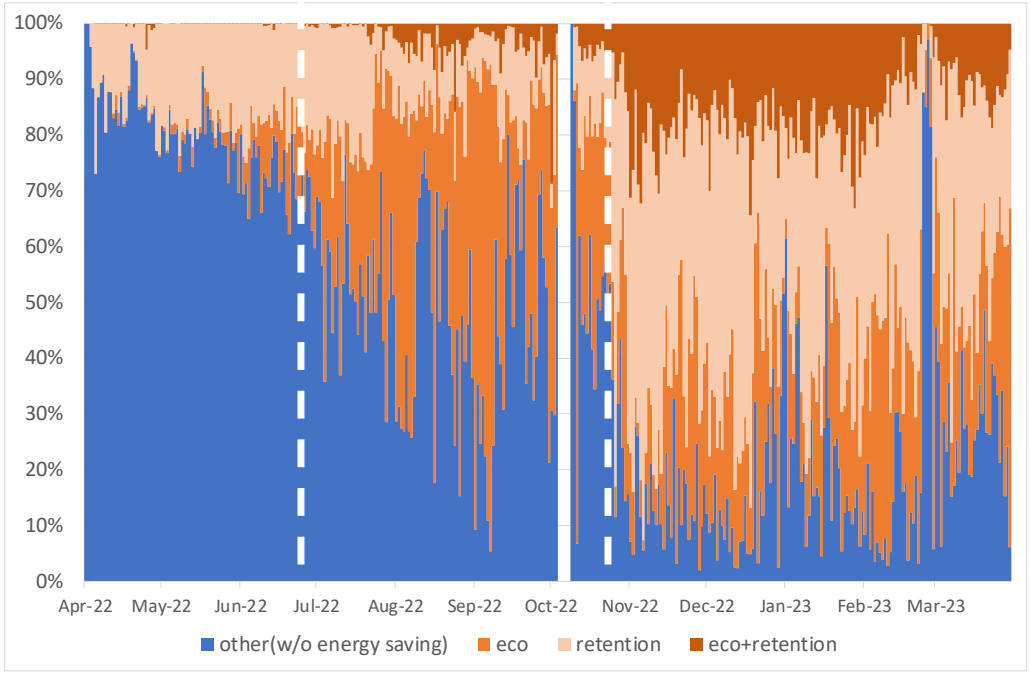


- Users can redeem the points to run their jobs with higher priority.



Result in FY2023 (April 2023 – March 2024)

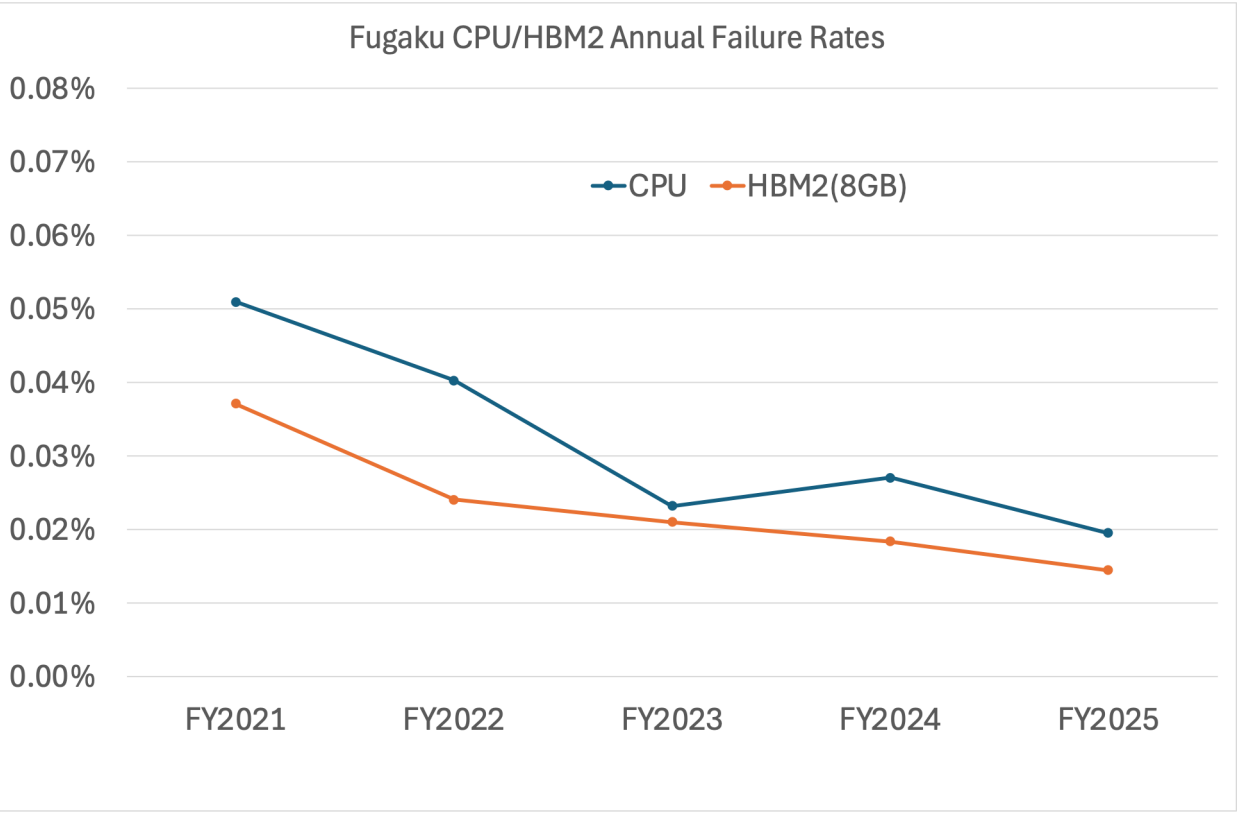
	FY2023 1H April-September	FY2023 2H October-March	+/-	
# of projects earned Fugaku point	**161**	**170**	+5.6%	
Average of Fugaku point per project	**248,303**	**320,692**	+29.2%	
Average power consumption per node(W)	**97.69**	**94.07**	-3.7%	-6.86% for FY2022

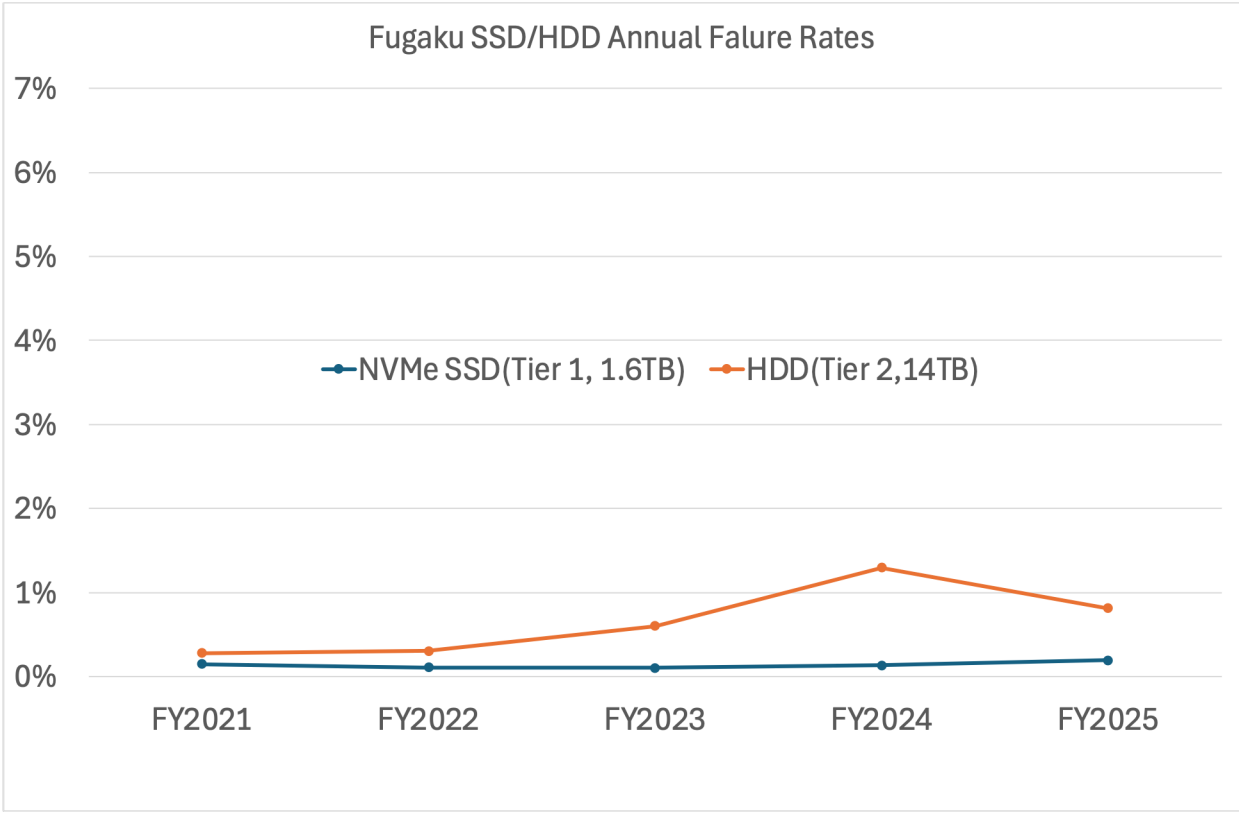


- Fugaku point program encouraged users' awareness for power saving.
- Average power consumption per node is significantly reduced.

Next step: Did the program contribute to reduce energy to solution?

Annual Failure Rates of Fugaku





- It looks it hasn‘t reached the bottom of the bathtub curve yet
- The rates are general and acceptable. It is rising but has not yet reached a critical level.

Lessons Learned

- **Hardware**
  - CPU, Memory seems to be OK, but storage must be serious
    - A procurement and operational plan to replace storage every five years is realistic
  - Secure replacement parts before they are discontinued
    - ex. Optical cables
- **Software**
  - A careful OS upgrade plan consistent with roadmap of the distributions
    - OS must be kept newer for security vulnerabilities and tool updated frequently (ex. PyTorch -> MPI/UCX(,CUDA) -> OFED(,GPU driver) -> Linux kernel)
    - Does not rely on software that would prevent OS upgrades
- **The operation period should be fixed in advance**